ERDC TN-DOER-C15
July 2000

Determining Recovery Potential of Dredged Material for Beneficial Use – Site Characterization: Statistical Approach

PURPOSE: This technical note is the third in a series of three technical notes providing guidance on evaluating the potential for recovery of dredged material for beneficial use (BU), either as is or using physical separation (soil washing) to meet BU specifications. This technical note introduces statistical methods for developing a sampling plan and interpreting and extrapolating the resulting data. The first technical note (Olin-Estes and Palermo 2000b) introduces physical separation concepts and presents mathematical relationships for estimating material recovery potential (MRP). A prescriptive approach to estimating volumes meeting BU requirements based on available information, or information obtained from limited sampling, is outlined in the second technical note (Olin-Estes and Palermo 2000a).

BACKGROUND: The principal motivation for BU recovery of dredged material is the growing shortage of storage capacity in confined disposal facilities (CDFs). The fundamental purpose of these technical notes is to assist in determining when material recovery is technically and economically feasible, and provide a strategy for obtaining and using physical and chemical information necessary for this evaluation at the least possible cost. The fundamental approach is to begin with available information and progress to targeted sampling and analysis as needed.

Olin-Estes and Palermo (2000a, 2000b) introduce prescriptive (limited sampling) site characterization methods, and physical separation concepts and methods for estimating MRP, respectively. The feasibility of separation as a management approach is dependent on several factors, including ability to identify distinct fractions within the material meeting BU criteria, ability to separate suitable fractions, and MRP as determined by available volumes of suitable material. This technical note introduces statistical sampling and data estimation methods for extensive site characterization.

INTRODUCTION: When separation appears to be necessary to meet material specifications for identified BUs, more detailed sediment/site characterization and evaluation are needed to estimate MRP. Extensive site sampling and data interpretation are addressed in the following sections. Figure 1 illustrates the position of extensive site sampling and characterization in evaluating feasibility of BU recovery. Olin-Estes and Palermo (2000b) describes the overall evaluation approach more extensively.

DATA REQUIREMENTS: The objective of extensive site characterization is to address the same data requirements as described in the two previous technical notes. These requirements are repeated here for clarity and ease of reference.

There are essentially two levels of MRP estimates: screening level, based on existing information, and definitive, based on more extensive site sampling. Several types of data are required to estimate MRP:

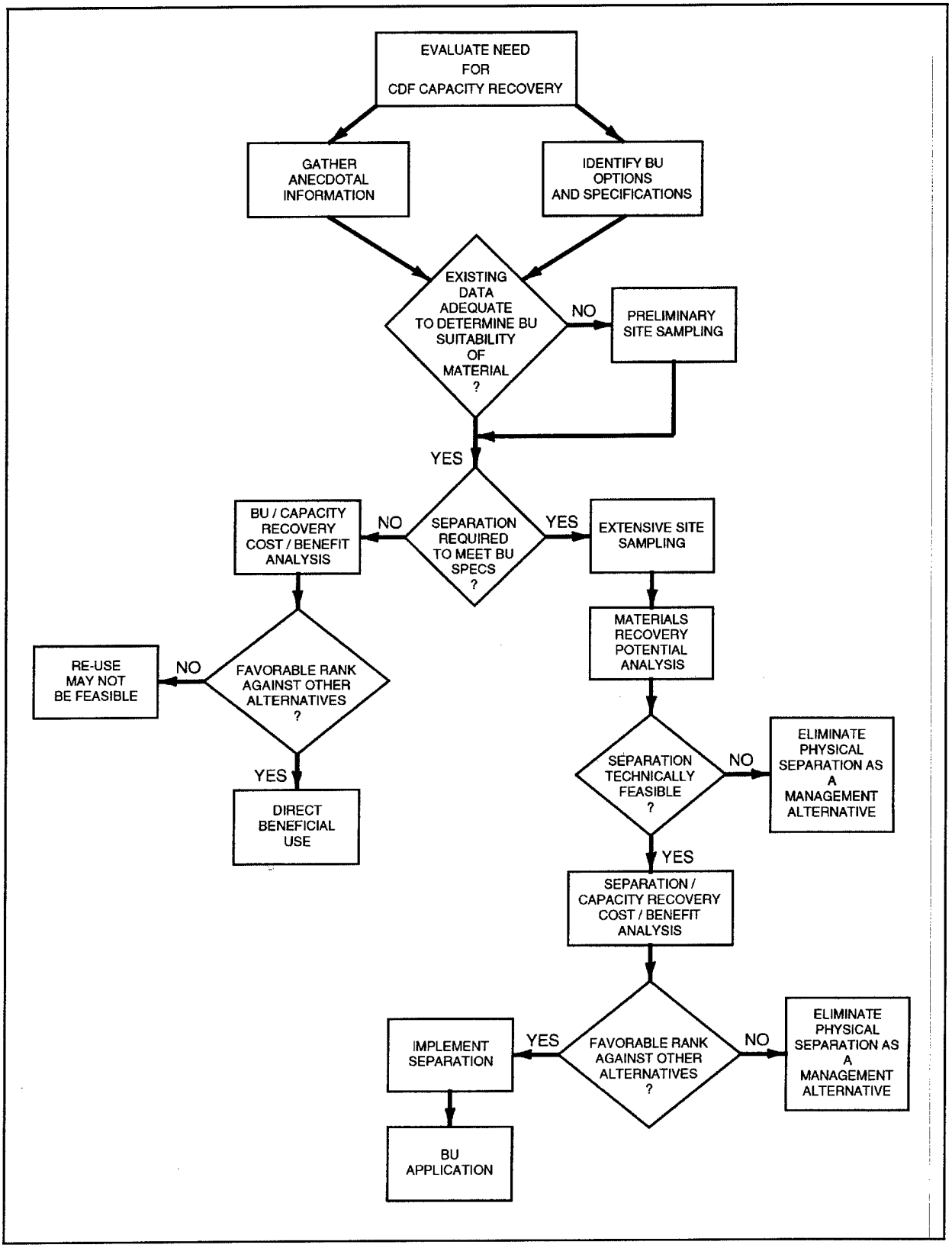

Figure 1. Evaluation of feasibility for BU recovery of dredged material

- Bulk sediment data:
  - Volume of available bulk sediment or dredged material.
  - Grain size distribution (GSD) of the bulk material (prior to separation).
  - Concentrations of contaminants of concern (COC) in the bulk sediments.
- BU specifications, including acceptable GSD and COC levels.
- Concentrations of COC in material fractions, if separation is determined to be necessary to meet BU specifications.

Use of existing information to obtain screening level estimates of MRP is described in Olin-Estes and Palermo (2000a). This technical note addresses the case in which existing information is inadequate for definitive determination of BU feasibility and MRP. Because data are rarely available for in-CDF materials, and because physical and chemical data for in-channel materials are not generally obtained specifically for determination of BU potential or physical separation feasibility, most projects of any size will ultimately require an extensive sampling effort.

General sampling considerations (sampling methods and equipment, sample volume requirements, analyte selection, depth of sampling, sample replication and compositing, and physical testing) are the same for both the statistical and prescriptive site characterization approaches. These are discussed fully in Olin-Estes and Palermo (2000a). This technical note address specifically the statistical basis and procedures for developing a site sampling plan and interpreting and extrapolating data.

SITE CHARACTERIZATION USING STATISTICAL APPROACHES: In designing a sampling plan, in addition to using available information about the site, it is often helpful to look at the tools available for interpretation of the resulting data. A number of statistically based approaches provide tools for determining the number of samples required to determine a measured parameter with a specified degree of confidence, unbiased approaches for structuring a sampling plan, and methods for interpreting and extrapolating data (Winkels and Stein 1997; Keillor 1995; Keillor 1993; Lubin, Williams, and Lin 1995; Isaaks and Srivastava 1989). Given the constraints of time and budget, the number of samples required based on statistical considerations will often be much larger than is physically or economically feasible to obtain or analyze, unless the variability of the material is quite low. However, a sampling plan certainly should not be implemented without considering a statistical design, even though modifications to that design may ultimately be required. The resulting data will then lend itself to statistical analysis and available methods for extending the data to unsampled areas. Appendix I presents a glossary of statistical terms used in the following discussion.

Statistical Analysis. In general, the larger a data set is, the more it tends toward a normal distribution. This is important because when it can be established that data are normally distributed, there are a number of statistical tools to help interpret the significance of differences between samples and to predict the likelihood of values falling outside a specified range. Among these are the t-test, the Paired Difference Test, and Analysis of Variance (ANOVA). However, most environmental data are not normally distributed. Due to cost constraints, the data sets are too small, or may contain many zero values due to the heterogeneity of deposits and the difficulty of obtaining representative samples. Because environmental data do not always meet the requirements and

assumptions for standard (parametric) statistical methods, nonparametric methods are sometimes useful. Nonparametric methods use the ranking of the data values, rather than the individual data values themselves. No assumptions regarding the distribution of the data are required for nonparametric methods (Mendenhall and Beaver 1994). Several of these methods do require a minimum number of samples to be applicable, and these requirements should be reviewed during the sampling planning stages. Among them are the Mann-Whitney test for comparison of the means and variances of two independent samples; the Sign Test for Paired Observations, which can be used to determine if values of a selected parameter are greater in one sample than in another (the nonparametric paired t-test, which has a binomial distribution under certain conditions); and the Kruskal-Wallis H-test, which is used for determining whether multiple samples come from the same population (the nonparametric analysis of variance test, which has a chi-square distribution under certain conditions).

The primary utility of parametric and nonparametric methods is to determine if there is a statistically significant difference between samples or sample means. These tools may be useful in interpreting the data and extending it to unsampled areas. Before getting to that point, however, a sampling plan that will produce data lending itself to statistical analysis must be developed. The key questions in developing a sampling plan are where to sample, how many samples to take, what size samples are required, and what parameters to analyze. The first two questions can be addressed statistically. The latter two are addressed in Olin-Estes and Palermo (2000a).

Developing a Sampling Plan Using Statistical Methods. Statistical packages have been developed to assist in design of sampling plans and/or identification of hot spots that could be adapted to determine the number and location of samples required to characterize a CDF. The STATSS (Statistical Techniques Applied to Sediment Sampling), a guidance document prepared for the U.S. Environmental Protection Agency, Region 5 (Lubin, Williams, and Lin 1995), describes statistical considerations of sampling, and approaches for determining grid and sample size for sampling sediments within a waterway. The Groundwater Modeling System (GMS) (Brigham Young University 1999) is another statistically based package designed to facilitate definition of subsurface contaminant plumes. The following is a general discussion of the underlying statistical principles and data analysis methods that provide the framework for statistically based sampling. The reader is referred to these statistical packages and references for more in-depth guidance in applying these principles.

Where to sample. There are three basic sampling approaches:[1]

- Judgmental approach.
- Random approach.
- Systematic approach.

A judgmental approach involves applying what is known about a site, and sampling in those areas that appear most likely to be contaminated or otherwise of interest. The judgmental approach is

[1] Personal communication, 9 October 1998, Dr. John H. Pardue, Civil and Environmental Engineering Department, Louisiana State University, Baton Rouge.

essentially the prescriptive approach described in Olin-Estes and Palermo (2000a). A systematic approach involves imposing a uniform grid over the area of interest and sampling from the same location in each grid. A random approach involves selecting sampling points within a gridded area using a random number generator to choose from among the alternative sample locations. The random approach is optimum from a statistical standpoint but in environmental sampling may not be the best choice. If the number of samples being taken over a large area is small, purely random sampling could well miss an area of known contamination. The judgmental and systematic approaches help to compensate for this, but may violate the assumption of randomness required in statistical analysis. In practice, the following three principal sampling approaches that are used in environmental sampling incorporate a combination of these elements:

- Systematic random.
- Judgmental random.
- Systematic judgmental.

As previously mentioned, a uniform grid is imposed over the area to be sampled in the systematic methods. In the systematic random method, a random number generator is then used to pick the locations within the grid that are to be sampled (Figure 2). Alternatively, one may select only every n^{th} member from the sampling grid with the starting element randomly selected (Lubin, Williams, and Lin 1995). Systematic judgmental methods focus attention on the area most likely to be contaminated, a grid is imposed, and a sample is taken from the center of each grid (Figure 3). The judgmental random method involves separating the area of interest into blocks that are expected to contain similar samples (such as similar levels of contaminants). A grid is imposed over these areas and sample sites are selected randomly within each block (Figure 4). This is also referred

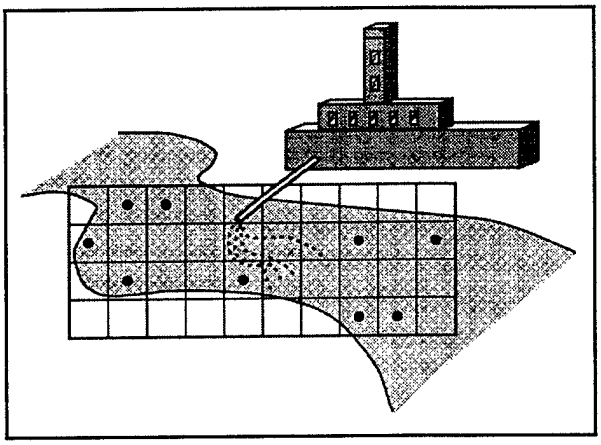

Figure 2. Systematic random sampling

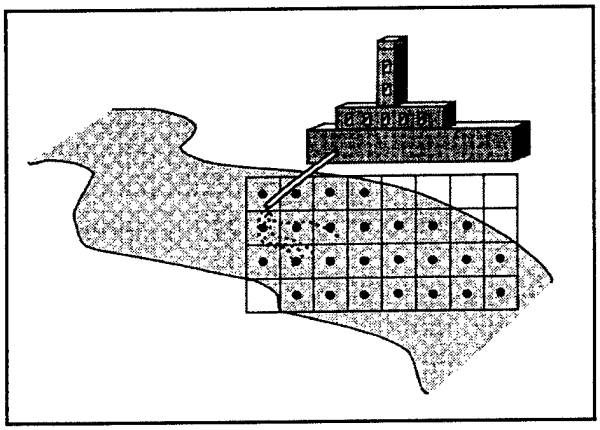

Figure 3. Systematic judgmental sampling

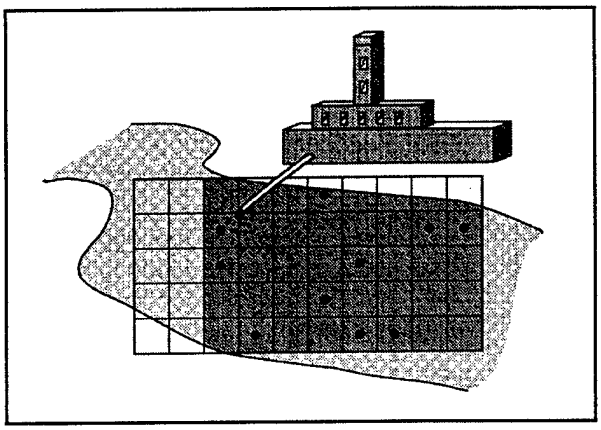

Figure 4. Judgemental random sampling

to as stratified random sampling (Lubin, Williams, and Lin 1995). The sampling method will ultimately be selected based on the greatest confidence in capturing representative data, quality and availability of existing information on which to base the method selection, and cost considerations. Additional discussion can be found in U.S. Environmental Protection Agency/U.S. Army Corps of Engineers (1995).

Estimating the number of samples required. Ultimately, the number of samples obtained will be determined by cost considerations. The upper threshold will almost certainly be set by the number of samples required to determine the desired parameter (e.g., contaminant concentrations, percent sand) with a specified degree of confidence. If a normally distributed sample can be assumed, then from the empirical rule, approximately 95 percent of the values will lie within 1.96 s of the mean, where s is the standard deviation of the sample. An acceptable margin of error can then be used to estimate the number of samples required. For example, to calculate the mean concentration of a constituent at a selected depth within 10 mg/kg at the 95 percent confidence level, then:

$$1.96\frac{s}{\sqrt{n}} = 10 \tag{1}$$

Solving for n gives the number of samples required to determine the mean within 10 mg/kg, at the 95 percent confidence level. Higher or lower confidence levels can be used. Further discussion can be found in Mendenhall and Beaver (1994). The obvious disadvantage to this method is that some idea of the variability of the data to be obtained is required prior to sampling. One could use results from analysis of selected samples taken within the CDF to estimate s and determine how many additional samples should be analyzed. (The standard deviation for the subsample can be calculated directly, or the range of the data can be used to estimate s (Appendix I).) If no data are available, an action level can be used as an estimated value for the variance. Such an iterative approach is described by Lubin, Williams, and Lin (1995) using a mathematical relation for estimating sample numbers that does not use the mean, but does incorporate acceptable error levels (α and β). However, environmental data are typically highly variable (large s), which may result in unrealistically high numbers of samples required. Additionally, these approaches require the assumption of a normal distribution, which is not typical of most environmental data. The geometric alternative variance can be used to estimate required sample size for lognormally distributed data; this approach is further described in Lubin, Williams and Lin (1995). Another alternative is to sample sequentially, evaluating data as they are generated and continuing to sample until a definitive threshold is achieved at a desired confidence level. The sequential approach and additional methods for estimating required sample numbers for different grid configurations and confidence levels are described in Lubin, Williams, and Lin (1995).

Several of the nonparametric data analysis methods require a minimum number of samples and observations to be valid, or require equally paired numbers of observations between samples to be compared. For example, the Kruskal-Wallis H-test (nonparametric ANOVA) requires at least three samples with at least three observations per sample. When there are more than 6 observations per sample, the distribution of the H statistic is well approximated by the chi-square distribution (McBean and Rovers 1998). The STATSS (Lubin, Williams, and Lin 1995) guidance document provides simple guidance for determining the number of samples required for a specified error level or confidence interval.

Sample size required. This is distinct from statistical sample size; in this instance sample size refers to the volume of material that is homogenized and then sampled for analysis. For example, if a 1.8-m (6-ft) core is taken, it will normally be subdivided into smaller sections that are thoroughly homogenized. Then a very small subsample of each homogenized section is taken for chemical analysis. Because sediments and the distribution of contaminants within the sediments are typically very heterogeneous, homogenization volume is a relatively important factor in obtaining data that are representative of site conditions. Additional information regarding the influence of sample size and replication in capturing the effects of material heterogeneity is found in Olin-Estes and Palermo (2000a).

Interpreting and Extrapolating (Estimating) Data. Examining the different ways in which available data can be grouped and manipulated to reveal trends may be one of the most practical approaches to determining where to sample and how many samples to take. Isaaks and Srivastava (1989) present a clear discussion of a number of methods for grouping data and extrapolating existing data to unsampled points, specifically directed at taking a practical approach to the application of statistical theory. Although many of these methods will be helpful in maximizing the information obtainable from a limited data set, the user should be aware that the results obtained from statistical analysis of the data may differ for different assumptions. Statistical analysis offers an improvement over "best guess" determinations of parameter distributions, but is not a foolproof method. One reason in particular is that the geostatistical methods described by Isaaks and Srivastava (1989) are based on the assumption that the values of interest are spatially continuous. This is probably a reasonable assumption for natural, undisturbed materials over limited areas. For disturbed materials, such as dredged material disposed in a CDF, this is a more difficult assumption to make. However, the distribution of hydraulically placed dredged material in a CDF is a result of natural processes (settling velocities), assuming the material has not been otherwise disturbed. Under these circumstances, continuity may be a reasonable assumption for limited areas of the CDF. For example, gradation of particle size and contaminant levels would be expected in moving from the inlet area to the outlet of a CDF in which the material is hydraulically placed; two or three distinct zones might be expected.

Interpreting data.

- *Univariate Data* – Data pertaining to a single variable can be presented very simply in a relative location map (Isaaks and Srivastava 1989). For example, if a uniform grid is imposed on the sampling area, and a sample taken from the center of each grid, the resulting value for the parameter of interest can then be superimposed on a map of the area, giving an indication of spatial distribution. A frequency histogram may also be used to give a quick visual on the predominantly occurring values. A cumulative frequency table will be useful in illustrating what percentage of samples fall below a certain threshold; this is a particularly useful technique where contaminant concentrations are of interest. Tests for normality or lognormality should be conducted as a matter of routine to establish whether or not the distribution of the data falls within either of these two categories. Typically, environmental data do not, but this should be done as a matter of practice. Summary statistics, including the mean, range, minimum, maximum and standard deviation, should be determined for the data, which may be grouped by zones if that provides a more meaningful result. The spatial distribution of values will suggest appropriate groupings, if any.

Parameters of interest in a CDF are likely to include percent sand, percent clay, and contaminant concentrations. The spatial distribution of each of these parameters can be examined individually, but by looking at the relationships between these parameters, it is likely that much can be determined about the material distribution within the CDF using physical parameters and more limited, targeted, chemical analysis. Bivariate data analysis methods provide the means to do this.

- *Bivariate data* – Bivariate data analysis methods permit the comparison of two parameter distributions to determine whether a functional relationship exists between them (Isaaks and Srivastava 1989). Likely to be of interest in determining the distribution of recoverable materials in a CDF is the relationship of percent sand and percent clay to contaminant levels. Summary statistics and tests for normality should be calculated for each distribution individually. A relative location map can be employed, as for the univariate data, giving the values of each parameter as a function of spatial distribution. A scatter plot of the two parameters, one plotted on the ordinate and the other on the abscissa, may illustrate any functional dependence that exists. The linearity of the relationship of the variables can be evaluated using the correlation coefficient ρ, defined in Appendix I. The correlation coefficient varies between -1 and +1; +1 indicates a straight line with a positive slope (positive correlation), -1 indicates a straight line with a negative slope (negative correlation), and values near zero indicate little or no correlation between the variables (Isaaks and Srivastava 1989). For example, one would expect particle size and contaminant concentration to be negatively correlated and percent clay and contaminant concentration to be positively correlated, contaminant level decreasing with increasing particle size. If the correlation coefficient is unduly influenced by a few extreme values, the rank correlation coefficient may be a more useful statistic. This is further described in Isaaks and Srivastava (1989).

- *Censored data* – In environmental sampling, a high percentage of samples may have no measurable contaminants (nondetects). Concentrations of these analytes, known as censored values, are normally reported as less than the method detection level (<MDL). The actual concentration of the contaminant lies somewhere in the range from zero to the MDL. There are several approaches to handling censored values. One approach is to ignore these values, which results in an overestimate of the mean and underestimate of the standard deviation (McBean and Rovers 1998). This alternative is acceptable only when the number of nondetects is very small. Alternatively, the censored values can be assumed to be equal to the detection limit, but this also introduces bias into the summary statistics. This alternative is preferred when the values are not highly variable and are near the MDL. A third alternative is to assume the censored values to be equal to MDL/2; this is the preferred alternative when the contaminant is present in highly variable concentrations. There are a number of statistical methods, parametric and nonparametric, for dealing with censored data; these are further described in McBean and Rovers (1998).

- *Spatial analysis* – Several variations of data groupings are possible based on the relative location map previously described. It may be visually instructive to identify the lowest and highest values on the map, or to replace individual data points with symbols based on assignment to certain ranges. An indicator map uses only two symbols, designating those data points falling above and below a specified threshold (Isaaks and Srivastava 1989). The

indicator map would likely be most useful for visualizing material in a CDF falling within a certain specification. However, a three-dimensional representation of the material in a CDF is needed. Different elevations within the CDF could be mapped separately, and vertical sections mapped in the same manner as the areal sections.

Another useful grouping tool is moving window statistics (Isaaks and Srivastava 1989). A uniform grid of data points is divided into subareas, and the mean and standard deviation of the parameters within each subarea are calculated and remapped at the center of the subarea. This results in a location map in which the parameter value trends and variability are easily seen. Overlapping the subareas can address the need for a sufficient number of data points within each subarea to provide reliable statistics (mean and standard deviation) while keeping areas small enough to capture local detail. Overlapping is particularly useful for small or irregular data sets (Isaaks and Srivastava 1989). Contour maps may also provide a useful visual description of material distribution, although their quantitative value may be limited where extensive interpolation is required. Plots of standard deviation versus sample means, h-scatter plots, correlation functions, covariance functions, and variograms are other available interpretive tools (Isaaks and Srivastava 1989) that might be considered if the basic summary statistics do not reveal a meaningful trend.

Estimating data. Estimating parameters for unsampled locations based on a limited data set is central to environmental characterization problems. A number of methods have been developed under the umbrella of geostatistics that have potential application. All are subject to the same inaccuracies as a result of site variability. Local estimates based on data that are highly variable are not likely to be very accurate, and should be interpreted in light of the confidence associated with the data set and the degree of spatial continuity evidenced by the data set.

The first step in estimating is to define the problem. The following three features of estimating a problem are adapted from Isaaks and Srivastava (1989):

- Is a global or local estimate desired?
- Is an estimate of the mean or the complete distribution of data values desired?
- Are point estimates or block values desired?

To characterize the deposits within a CDF, some point estimates will most likely be needed to identify extreme values (particularly with respect to contaminant levels), mean block values for particle size, and some estimate of the variability of the data to estimate recoverable volumes of material.

All of the methods discussed in Isaaks and Srivastava (1989) involve weighted linear combinations of the known data points:

$$\text{estimate} = \hat{v} = \sum_{i=1}^{n} w_i v_i \qquad (2)$$

where

$\hat{v}$ = the data point being estimated

w_i = a weighting factor

v_i = a data value

In estimating, adjustments are made to the sample weighting factor for distance from the point being estimated and clustering of data points. Samples closest to the data point being estimated will be given more weight than those at a greater distance. Data points that are clustered close together rather than uniformly distributed over an area will be given less weight because they are not representative of the larger area and may unduly influence the value of estimated global parameters. Isaaks and Srivastava (1989) describe a number of two- and three-dimensional declustering approaches. Additionally, closely spaced samples having similar values contain redundant information, and sample weights should be adjusted for this factor as well.

An important point to note is that distributions estimated from data points are volume dependent (Isaaks and Srivastava 1989); that is, if a homogenized 0.3-m (1-ft) section of core constitutes a single data point, the distribution estimated with this and like data points constitutes the distribution of parameter values for homogenized 0.3-m (1-ft) core sections. For this application, however, parameter values are desired for larger volumes of a scale that can be practically and economically excavated. Estimates of recoverable materials based on core analysis may not reflect the averaging that occurs when the material is excavated in larger volumes. Correcting for the error introduced by extrapolating small volume estimates to large volumes is a difficult problem, but some effort should be made to evaluate the potential effect of this factor. One approach might be to examine the standard deviation of the means of individual core sites.

Geostatistical estimating methods require identification of a model upon which the estimates are based (Isaaks and Srivastava 1989). A deterministic model can be used if enough is known about the process effects being measured to quantify them. For example, Stokes' law might be used to model the expected distribution of grain sizes across a portion of a CDF based on settling velocities, and extrapolate between data points to estimate the location of transitions across a certain grain size threshold. (Note that Stokes' law applies only to discrete settling of individual, nonflocculating particles. Discrete settling of fines does not normally occur in a CDF; thus the model would be applicable only to the coarse material in the CDF.) However, it is unlikely that most CDFs have been operated in a manner consistent enough to use this approach. Probabilistic models are used when no suitable deterministic model is available; in this approach the sample data are viewed as the result of some random process (Isaaks and Srivastava 1989). The most common parameters used in probabilistic approaches are the mean, or expected value, and the variance.

Global estimation is the determination of mean parameter values for large areas. Point estimation is the estimation of parameter values for small areas, or specific locations (Isaaks and Srivastava 1989). Declustering methods are used in both global and point estimation when samples are clustered rather than distributed over the entire area of interest. Point estimation methods also require weighting of sample values to reflect the relative distance from the point or area being estimated. Two point estimation methods, polygons and the local sample mean method, are

adaptations of inverse distance declustering methods. In the polygonal method, the sample value closest to the point being estimating is selected as the estimate. This value holds throughout the polygon of influence constructed around the estimated point. This method results in discontinuous parameter distributions over the area of interest (Figure 5).

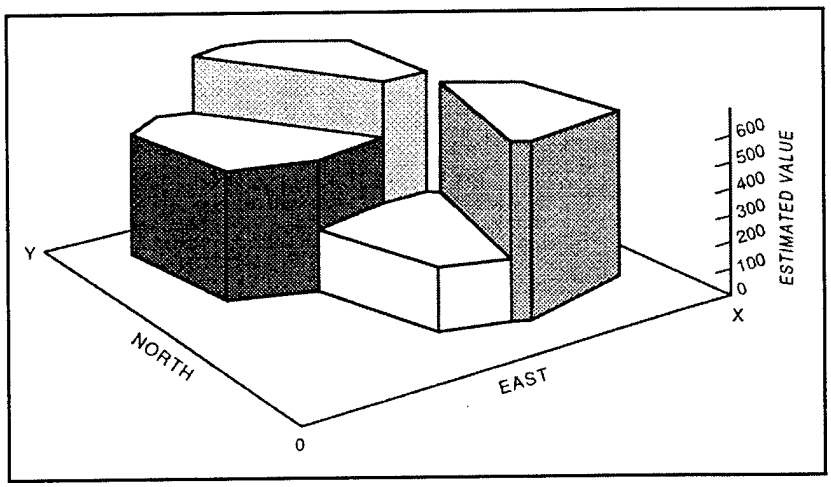

Figure 5. Polygonal point estimating method (adapted from Isaaks and Srivastava 1989)

The method of triangulation eliminates these discontinuities by fitting a plane through three samples surrounding the point to be estimated. An equation of the plane is developed that can be solved for the estimated parameter value at any point within the triangle by substituting the coordinates of the point. Alternatively, weighting factors can also be derived using triangulation. Inverse distance methods apply a weighting factor to nearby samples that is inversely proportional to the distance of the data point from the point being estimated. Some power p of the distance may also be used; small values of p decrease the difference in the weighting factors and larger values of p increase the difference. These methods are more fully described in Isaaks and Srivastava (1989).

Selection of nearby samples used as the basis for point estimation is also an important step in the estimating process, and may also be a consideration in location of initial sampling points. Isaaks and Srivastava (1989) refer to areas containing relevant samples as "search neighborhoods." Within the search neighborhood there must be a sufficient number of nearby samples, but not too many or redundant samples. The relevance of samples falling within the search neighborhood should also be considered. The number of samples to include is particularly important to inverse distance and kriging. The number of samples included using geometric estimating techniques is self-determining, based upon the orientation of the samples.

Normally, all available samples within the defined search neighborhood are used in estimation. Typically, an ellipse is centered on the point being estimated, with the long axis oriented in the direction of greatest continuity of the sample values (Isaaks and Srivastava 1989). In a CDF, this would likely be horizontally across the cell, perpendicular to the direction of flow. The length to width of the ellipse is determined by judgment, based on the degree of anisotropy evidenced in the available data.

Alternatively, all samples within a specified distance of the point to be estimated might be used. For regularly gridded data, the search neighborhood should be at least large enough to include the four nearest samples. In practice, a minimum of 12 samples is typical (Isaaks and Srivastava 1989). The search neighborhood for irregularly gridded data should be just larger than the average spacing between the sample data, estimated as follows (Isaaks and Srivastava 1989):

$$Average\ spacing\ between\ samples = \sqrt{\frac{Total\ area\ covered\ by\ samples}{Number\ of\ samples}} \quad (3)$$

At the same time that one must be concerned with having a sufficient number of samples for estimating, too many samples can be problematic. Computations for estimating procedures such as kriging become cumbersome with too many samples. This can be addressed by compositing samples outside the immediate area of the point being estimated. This procedure is further described in Isaaks and Srivastava (1989).

Ordinary kriging is an unbiased estimating method that is intended to minimize the mean residual m_R, or error, and the variance σ_R^2 of the errors. A probability model in which the bias and the error variance can be calculated is used, and nearby samples weighted to give $m_R = 0$ and minimize σ_R^2. The sample weights will change as unknown values are estimated (Isaaks and Srivastava 1989). The weighting matrix w is derived by multiplying two matrices (C and D) constructed from a selected random function model and parameters. The mathematical development of this procedure is somewhat complicated and the relationship to the physical problem not readily apparent. Simply described, the matrices are composed of the covariances between sample data points and the point being estimated. The D matrix "provides a weighting scheme similar to the inverse distance methods" (Isaaks and Srivastava 1989); the covariance between any sample and the point estimated decreases as the distance between them increases. The difference between the D matrix and inverse distance weights is that the covariances can be calculated from a larger family of functions, rather than being limited to a single form $|h|^{-p}$ (where h is the distance between the points and p is an arbitrarily selected exponent, as previously described). In effect, the kriging distance can be considered a statistical distance, rather than the geometric distance of the inverse distance methods (Isaaks and Srivastava 1989). The C matrix takes into account spatial continuity and redundancy, automatically providing an adjustment for clustering of data points. Ordinary kriging is therefore less adversely affected by sample clustering than other estimating methods, although it is computationally more difficult.

One characteristic of ordinary kriging is that, for selected functions, some of the sample weights may be negative, although the sum of the sample weights will always be 1, a necessary condition of unbiasedness. The result is that the procedure can yield estimates larger than the largest sample value and smaller than the smallest sample value. Since the data set is unlikely to contain the most extreme values, this is advantageous. A disadvantage is that negative estimates may also result. These may be arbitrarily set to zero when negative values do not make physical sense, as in the case of concentrations (Isaaks and Srivastava 1989). Selection of an appropriate model and model parameters requires fitting available data with a suitable function. This procedure and the result of varying the function parameters are extensively discussed in Isaaks and Srivastava (1989). Additionally, the random function model can be selected to reflect the degree of anisotropy of the site. Obviously, judgment and experience are requisite to using this procedure.

Other methods of point estimation can be found in Isaaks and Srivastava (1989). While ordinary kriging provides a method of obtaining point estimates, block kriging is a procedure for estimating an average value within a prescribed block. The previously described estimating methods use the spatial continuity of a single variable to provide estimates for unsampled points. Cokriging is a

method that uses information from secondary variables, which may be correlated with the primary variable, to improve estimates. An example of this might be the use of grain size as the secondary variable to improved estimates of contaminant concentration, the primary variable.

The assistance of a statistician will undoubtedly be helpful in designing a sampling plan that will produce data suitable for statistical analysis and estimating procedures. This note presents only a general discussion of the procedures and major considerations; a thorough familiarity with and understanding of the procedures by the practitioner is warranted.

COMPLETION OF BU AND SEPARATION FEASIBILITY EVALUATION: Once a reliable estimate of MRP has been developed, the information can be used in completing the evaluation of BU and separation feasibility. If recovery potential matches the requirements for the BU applications under consideration, and separation is required, appropriate operational methods or equipment for separation is selected. A cost analysis can then be performed and the final decision on separation feasibility made. Procedures for equipment selection and cost estimating are described in Olin et al. (1999). If separation is not required, a more straightforward cost benefit analysis can be conducted.

CONCLUSIONS: Development of a reuse plan for a CDF or dredging project will require a multistep approach incorporating existing data, practical and/or statistical sampling approaches, and identification of local BU opportunities and requirements. Little field verification is presently available regarding the efficacy of one sampling approach over another in characterizing the distribution of materials in a CDF. As further field experience is gained, refinements can likely be made that will result in an optimal approach and greater confidence in the results. Physical separation is only one of several approaches that can be taken to produce material suitable for various BUs. Separation should be evaluated together with other alternatives to determine the most suitable approach for a given site.

POINTS OF CONTACT: For additional information, contact the author, Trudy J. Olin-Estes (601-634-2125, *olint@wes.army.mil)* or the Program Manager of the Dredging Operations and Environmental Research Program, Dr. Robert M. Engler (601-634-3624, *englerr@wes.army.mil).* This technical note should be cited as follows:

> Olin-Estes, T. J. (2000). "Determining recovery potential of dredged material for beneficial use – Site characterization: Statistical approach," *DOER Technical Notes Collection* (ERDC TN-DOER-C15), U.S. Army Engineer Research and Development Center, Vicksburg, MS. *www.wes.army.mil/el/dots/doer*

REFERENCES

Brigham Young University. (1999). "The Department of Defense Groundwater Modeling System; v3.0 reference manual," Environmental Modeling Research Laboratory, Provo, UT.

Isaaks, E. H., and Srivastava, R. M. (1989). *An introduction to applied geostatistics.* Oxford University Press, New York.

Keillor, P. (1993). "The challenge of describing a contaminated sediment site," course notes from "Managing Contaminated Sediments," A Short Course of the Department of Engineering, Professional Development, University of Wisconsin-Madison, April 13-15, 1993.

Keillor, P. (1995). "An approach to describing site characterization for remediation of contaminated sediments," draft white paper (for further information, contact the author of this technical note).

Lubin, A. N., Williams, M. H., and Lin, J. C. (1995). "Statistical Techniques Applied to Sediment Sampling (STATSS)," U.S. EPA Region 5, Draft 03, November 27, 1995 (for further information, contact the author of this technical note).

McBean, E. A., and Rovers, F. A. (1998). *Statistical procedures for analysis of environmental monitoring data and risk assessment.* Prentice-Hall, Upper Saddle River, New Jersey.

Mendenhall, W., and Beaver, R. J. (1994). *Introduction to probability and statistics.* Duxbury Press, Belmont, CA.

Olin, T. J., Bailey, S. E., Mann, M. A., Lutes, C. C., Seward, C. A., and Singer, C. F. (1999). "Physical separation (soil washing) equipment for volume reduction of contaminated soils and sediments," EPA 905-R99-006, Assessment and Remediation of Contaminated Sediments Program, Great Lakes National Program Office, Chicago, IL.

Olin-Estes, T. J., and Palermo, M. R. (2000a). "Determining recovery potential of dredged material for beneficial use – Site characterization: Prescriptive approach," *DOER Technical Notes Collection* (ERDC TN-DOER-C14), U.S. Army Engineer Research and Development Center, Vicksburg, MS. *www.wes.army.mil/el/dots/doer*

Olin-Estes, T. J., and Palermo, M. R. (2000b). "Determining recovery potential of dredged material for beneficial use – Soil separation concepts," *DOER Technical Notes Collection* (ERDC TN-DOER-C13), U.S. Army Engineer Research and Development Center, Vicksburg, MS. *www.wes.army.mil/el/dots/doer*

U.S. Environmental Protection Agency. (1994). "Remediation guidance document," EPA 905-B94-003, Assessment and Remediation of Contaminated Sediments Program, Great Lakes National Program Office, Chicago, IL. *http://www.epa.gov/glnpo/arcs/EPA-905-B94-003/EPA-905-B94-003-toc.html*

U.S. Environmental Protection Agency/U.S. Army Corps of Engineers. (1995). "QA/QC guidance for sampling and analysis of sediments, water, and tissues for dredged material evaluations - Chemical evaluations," EPA-823-B-95-001, U.S. Environmental Protection Agency, Washington, DC.

Winkels, H. J., and Stein, A. (1997). "Optimal cost-effective sampling for monitoring and dredging of contaminated sediments," *J. Environ. Qual.* 26, 933-946.

APPENDIX I
STATISTICAL TERMS AND DEFINITIONS

Terminology – Several terms are fundamental to an understanding and use of statistics in structuring a sampling plan and interpreting the resulting data. These are described briefly here. More rigorous definitions can be found in any text on statistics.

Sample – In statistical terms, sample refers to a group of observations taken from an overall population (as distinct from the common usage, which refers to a discrete amount of material that, when measured for certain parameters of interest, would compose one of the observations of a statistical sample). For example, the percent sand for each 0.3-m (1-ft) increment of a 1.8-m (6-ft) core could collectively be considered a sample. Various statistical parameters of this sample could be compared with those of other cores to determine whether apparent differences are greater than that which would be expected from the random variability of the data. If the samples are significantly different, this may be an indication of a trend, such as increasing or decreasing particle size as a function of location in the CDF.

Distribution – This refers to the shape of the graph resulting when the values of a data set are graphed against the number of times they occur. The most familiar distribution is the normal distribution, also known as the Gaussian, or bell-shaped, distribution (Figure I1). There are various tests for normality. The distribution of the data, if known, can be used to determine the probability of occurrence of a specific parameter value, such as concentration or grain size. Most environmental data are not normally distributed. Skewed distributions, with a long tail to the right or to the left, are common.

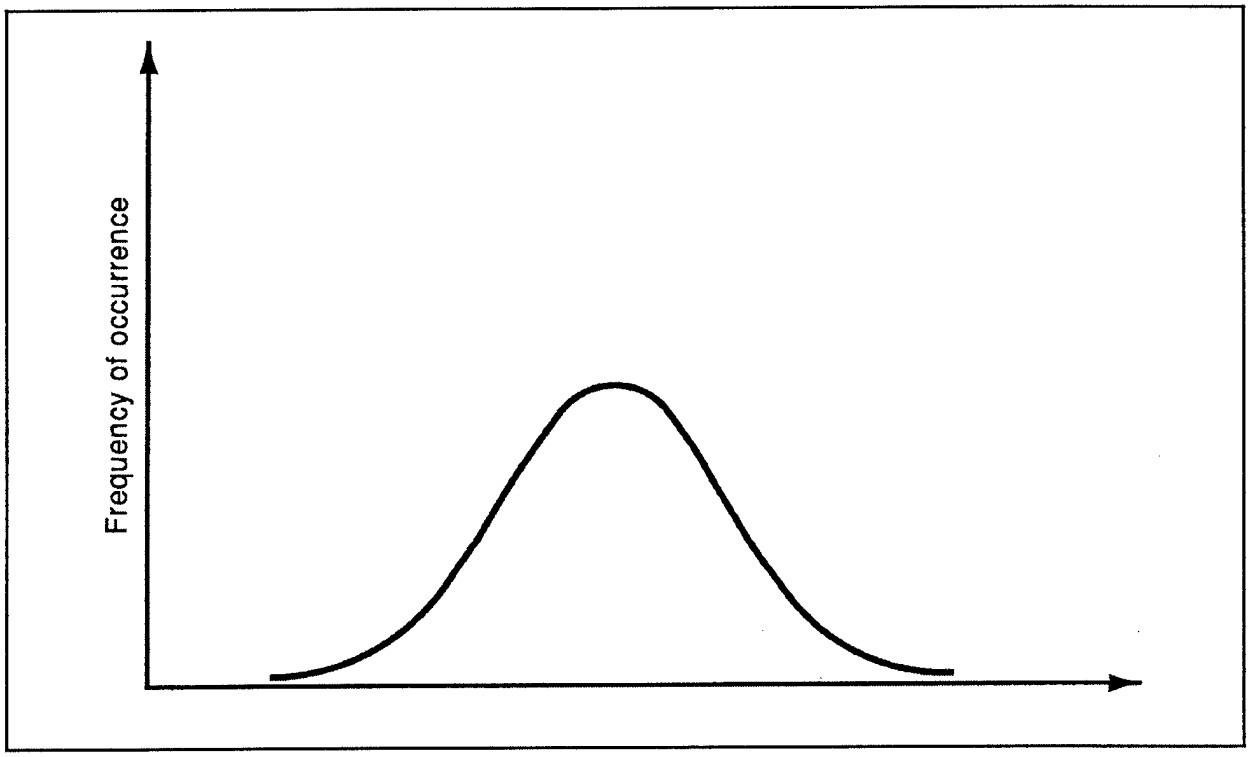

Figure I1. Normal (Gaussian) distribution

Mean – The mean is a measure of central tendency of the data, that is, the central value about which the data are grouped. There are several different types of means, including the arithmetic mean, the harmonic mean, and the geometric mean. The arithmetic mean $\bar{x}$ is most commonly used, but the harmonic mean H and geometric mean G are important when the data set includes a few very high or low values that may influence the arithmetic mean (McBean and Rovers 1998). The arithmetic mean is defined as follows, and has the same units as the individual data observations (e.g., mg/kg):

$$\bar{x} = \frac{\sum_{i-1}^{n} x_i}{n} \tag{I1}$$

where

x_i = individual observations (values)

n = the number of observations

The relationship between the different means is as follows:

$$H \leq G \leq \bar{x} \tag{I2}$$

Median – The median is also a measure of central tendency, and may be more reflective of the center of gravity of a skewed distribution than the mean (Figure I2). The median is found by ranking the n measurements from smallest to largest. If n is odd, the median is the value with rank $(n + 1)/2$; if n is even, the median is the value halfway between the measurements with rank $n/2$ and $n/2 + 1$, that is, the average of the two middle values (Mendenhall and Beaver 1994).

Mode – The mode is the most frequently occurring value of the measured variable.

Standard deviation – The standard deviation s is a measure of the scatter of the data (how closely the data are grouped around the central value, the mean). A small standard deviation indicates closely grouped data with little variability. A large standard deviation indicates data that are widely variable. The number of standard deviations a value is away from the mean $\bar{x}$ is an indication of its probability of occurrence. For example, by the empirical rule, for a data distribution that is approximately bell shaped (normally distributed, or nearly so), 68 percent of the values will be within one standard deviation of the mean ($\bar{x}$ plus n minus s), 95 percent of the values will be within two standard deviations of the mean, and most or all of the values will be within three standard deviations of the mean (Figure I3) (Mendenhall and Beaver 1994; McBean and Rovers 1998). A value falling more than three standard deviations from the mean has a very small probability of occurrence. Values within one to two standard deviations would be reasonably expected to occur. This is the predictive value of statistical application to sampling; it allows the data to be extended to speculate on values expected in unsampled areas. Data with low variability increase the level of confidence in determining how likely a certain value is to occur, or threshold to be exceeded.

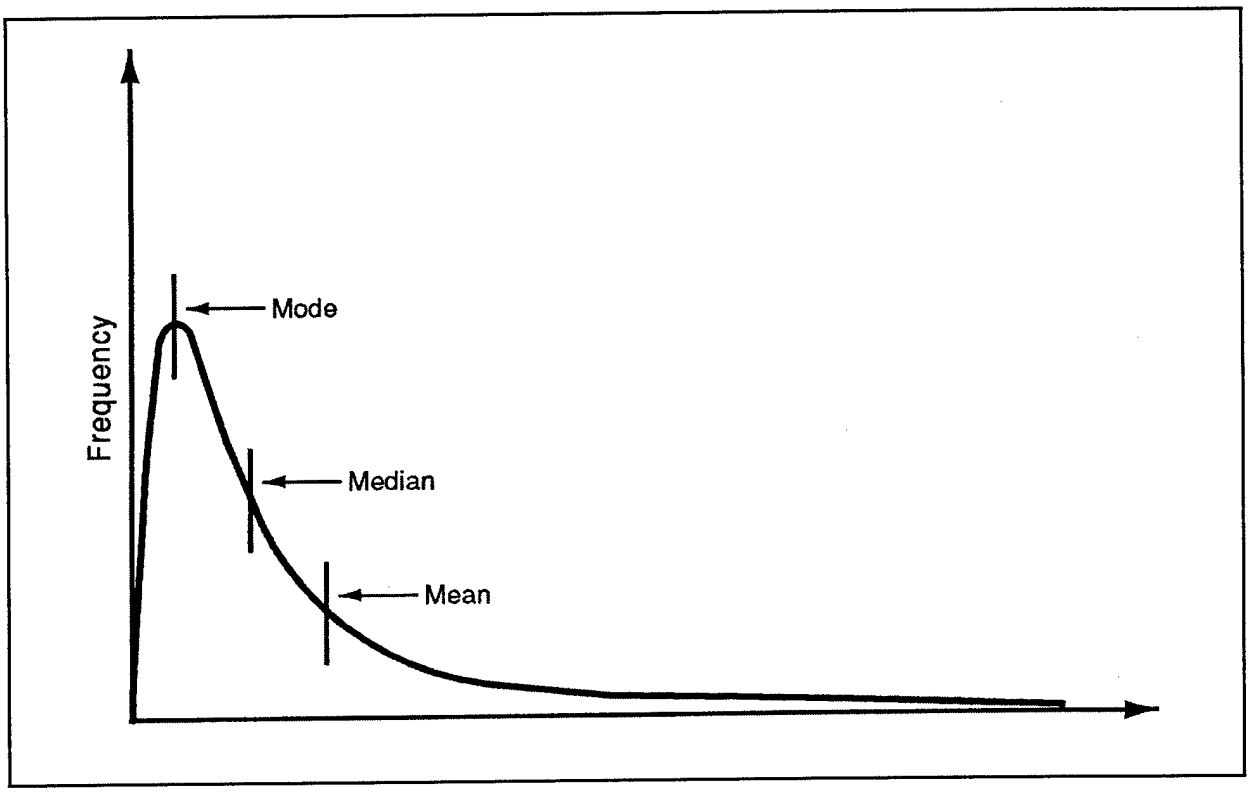

Figure I2. Mean, median, and mode

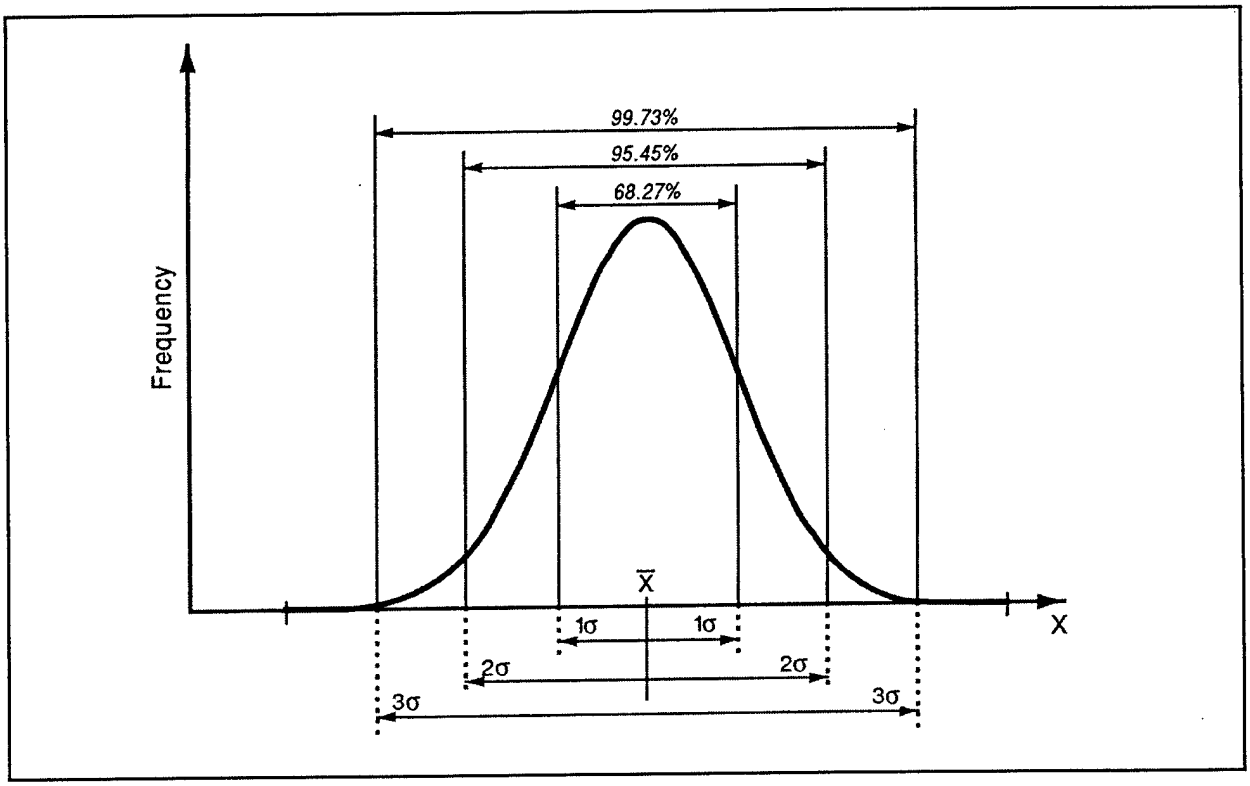

Figure I3. Standard deviations from the mean (adapted from McBean and Rovers 1998)

The standard deviation s has the same units as the mean and is defined as follows:

$$s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}} \tag{I3}$$

(The square of the standard deviation s^2 is known as the variance. The term is less commonly used because the units are squared and it is not, therefore, as intuitively useful; but it refers to the same characteristic of the data. A high variance indicates widely variable data; a low variance indicates little variability.)

Range – The range is also a measure of the scatter of the data, is useful for bracketing the extremes of the data, and can also be used for a rough estimate of the standard deviation ($s \approx R/4$). The range is simply the difference between the largest and the smallest value in the data set. The range can be influenced by the existence of a single extreme value. The 10-90 percent range is less sensitive to the presence of a few extreme data points, and is the difference between the highest and lowest values of 80 percent of the data. The interquartile range is a similar statistic, excluding the upper and lower 25 percent of the data.

Correlation coefficient – a measure of the linearity of the relationship between two parameters. The correlation coefficient is given as:

$$\rho = \frac{\frac{1}{n}\sum (x_i - \bar{x})(y_i - \bar{y})}{\sigma_x \sigma_y} \tag{I4}$$

where σ_i is the standard deviation of the population, estimated for a sample as s, as previously defined.

Covariance – The covariance is defined by the numerator of the correlation coefficient, and is sometimes used alone as a summary statistic. Dividing the covariance by the standard deviations ensures that the correlation coefficient is always between ±1, and is therefore independent of the magnitude of the data (Isaaks and Srivastava 1989).

Bias – "An estimator of a parameter is said to be unbiased if the mean of its distribution is equal to the true value of the parameter. Otherwise, the estimator is said to be biased" (Mendenhall and Beaver 1994).

Accuracy and precision – These are two terms of importance that are frequently confused. Accuracy refers to the closeness of a measured value to the true value. Measured values may differ from the true value as a result of instrument variability, operator error, losses in handling, and sample contamination from other sources. Precision refers to the repeatability of a method. Data that are accurate and precise are desirable, but accuracy cannot be determined for an unknown quantity (such as the concentration of lead in a soil sample). Data that are precise at least assure that methods

are consistent, and the error for a specific analytical method can be quantified using standards. Precise data should still be interpreted in light of identifiable factors that might give results different from the true value (inaccurate results). For example, concentrations that seem unreasonably high given what is known about the source of a material may suggest sample or instrument contamination from other sources. Concentrations that are unexpectedly low may suggest loss mechanisms, such as volatilization or bacterial degradation occurring in samples improperly stored, or other analytical error. Careful sample handling, adequate replication, and suitable quality control measures help to minimize these types of errors.